Beyond Frontiers

Scene-Anomaly Guided Autonomous Exploration

Akash Kumbar · Abhinav Raundhal · Madhava Krishna | Robotics Research Center, IIIT Hyderabad

*Autonomous exploration reframed as **geometric anomaly minimization** over learned indoor structural priors*

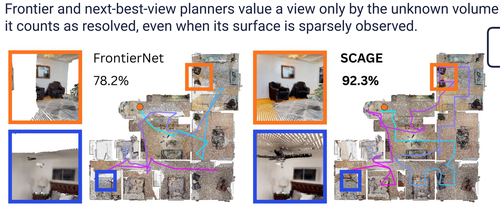
beyondfrontiers.github.io

IROS 2026 PITTSBURGH

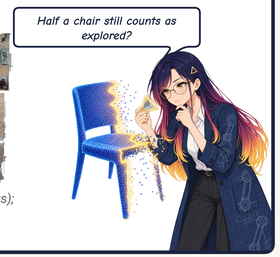
Motivation

Volumetric coverage ≠ Map fidelity

Frontier and next-best-view planners value a view only by the unknown volume it reveals; once a voxel is occupied it counts as resolved, even when its surface is sparsely observed.



HM3D scene 824: FrontierNet leaves the fan and chair only partially mapped (insets); SCAGE captures both completely.



Key idea. Learn what complete indoor geometry looks like, and explore where the map contradicts it.

Learning Indoor Structural Priors (Δ-Net)

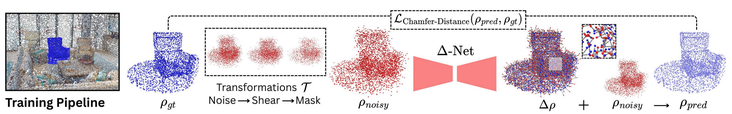
1 **Local chunking.** Structurally complete Matterport3D scenes are split into local chunks $\rho_{gt} \in \mathbb{R}^{N\times 3}$ (N = 4096), so the network learns recurring structural primitives.

2 **Stochastic corruption.** $\rho_{noisy} = \mathcal{T}(\rho_{gt})$: Gaussian noise, anisotropic shear and random masking emulate depth noise, distortion and occlusion.

3 **Denoising autoencoder.** A PTv3 encoder–decoder, trained from scratch, predicts a point-wise shift $\Delta\rho \in \mathbb{R}^{N\times 3}$:

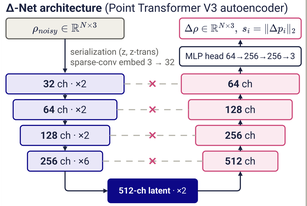
$$\rho_{pred} = \rho_{noisy} + \Delta\rho$$

4 **Anomaly score.** At inference, the correction magnitude $\|\Delta p_i\|$ measures how strongly each observed point violates the learned prior.



Training: ρ_{gt} is corrupted by $\mathcal{T}$; Δ-Net predicts the shift $\Delta\rho$ that restores ρ_{pred}, supervised by the Chamfer distance.

Δ-Net architecture (Point Transformer V3 autoencoder)



Encoder (indigo) / decoder (pink) PTv3 stages with ×2 grid pooling; ×n = blocks per stage; ✕ = no skip connection.

Objective. Bidirectional Chamfer distance:

$$\mathcal{L}_{CD} = \underbrace{\sum_{x\in\rho_{pred}} \min_{y\in\rho_{gt}} \|x-y\|_2^2}_{\text{accuracy: predictions lie on the true surface}} + \underbrace{\sum_{y\in\rho_{gt}} \min_{x\in\rho_{pred}} \|x-y\|_2^2}_{\text{completeness: the true surface is covered}}$$

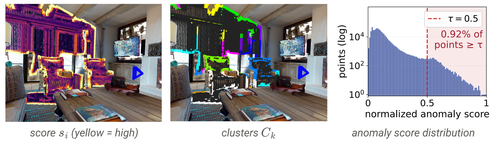
No skip connections. All information passes through the 512-channel latent, so the prior must be internalised rather than copied.

1 Observation & fusion

Depth is back-projected and fused into P, voxel-downsampled (5 cm) and split into overlapping 4096-point chunks (FPS + k-NN).

Exploration as geometric anomaly minimization

2 Scene-anomaly detection

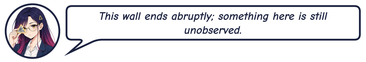


score s_i (yellow = high) · *clusters C_k* · *anomaly score distribution*

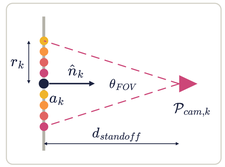
Δ-Net predicts a deviation Δp_i for every point; its normalised magnitude is the anomaly score:

$$s_i = \frac{\|\Delta p_i\|_2 - \min_j \|\Delta p_j\|_2}{\max_j \|\Delta p_j\|_2 - \min_j \|\Delta p_j\|_2} \in [0,1]$$

Points with $s_i > \tau$ (τ = 0.5) form the anomalous set $\mathcal{A}$, grouped into target clusters C_k by ball query.



3 Viewpoint generation & selection



standoff distance

candidate 6-DoF views

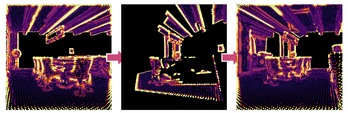
Each cluster C_k receives a 6-DoF view along the surface normal $\hat{n}_k$, at the distance that keeps it in view:

$$d_{standoff} = \max\left(d_{min}, \underbrace{\frac{r_k}{\tan(\theta_{FOV}/2)}}_{\text{cluster spans the FOV}}\right)$$

Candidates passing free-space and line-of-sight checks are ranked by:

$$u(T_k) = \frac{\overbrace{g(T_k)}^{\text{information gain}} \cdot \overbrace{\bar{s}_k}^{\text{mean anomaly}}}{\underbrace{\max(\|\mathcal{P}_{robot} - \mathcal{P}_{cam,k}\|_2, \epsilon)}_{\text{traversal cost}}}$$

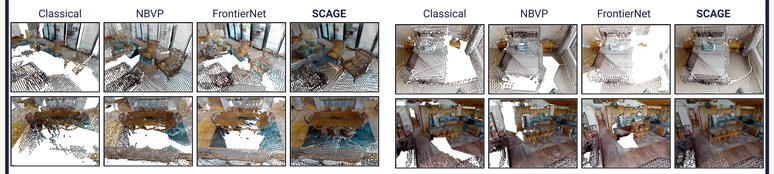
4 Navigation & resolution



view 1 · *view 2, before capture* · *view 2, after capture*

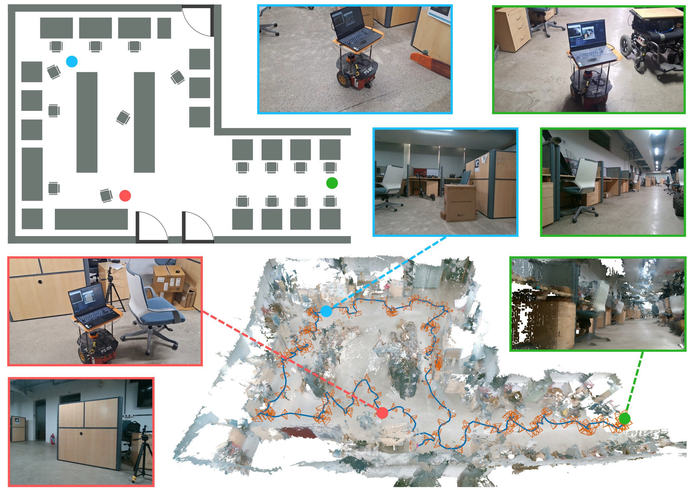
The robot reaches the selected pose via an RRT* path and fuses the new depth; the targeted anomalies are resolved.

Map Fidelity



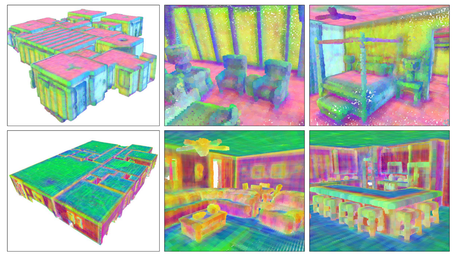
White = never observed. SCAGE captures cluttered objects (chairs, tables, beds) and broad planar surfaces alike.

Real-World Deployment



Pioneer P3-DX with a RealSense D455, ~7 Hz on a laptop GPU. Trained only on Matterport3D, Δ-Net flags incomplete chairs, cubicle partitions and walls in our lab.

Outlook: Emergent Structural Understanding



PCA of Δ-Net encoder features (HM3D 876, 879); no semantic labels used.

Structure without labels. Floors, ceilings and beds share features, while fans, chairs and cabinets stand apart.

Future work. Use the pretrained encoder as a structural backbone for semantic mapping, language-guided mapping and object-goal navigation.

Quantitative Results

89.7%

mean volumetric coverage ↑
best baseline (FrontierNet [8]): 72.3%

0.91

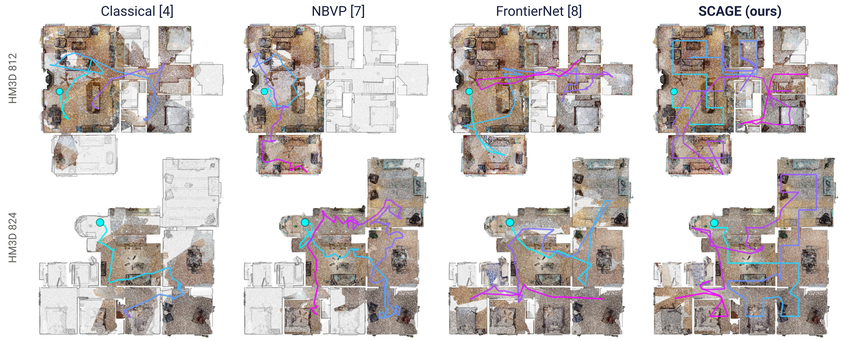
mean Chamfer distance (×10^{-3}) ↓
best baseline (FrontierNet [8]): 3.02

	Classical [4]	NBVP [7]	FrontierNet [8]	SCAGE
Coverage (%) ↑	50.7	62.8	72.3	89.7
Chamfer (×10^{-3}) ↓	6.55	4.60	3.02	0.91

With monocular depth (Metric3D v2 [32]): **88.1%** coverage, **1.89** Chamfer.

10 unseen HM3D validation scenes · FrontierNet evaluation protocol and step budgets · Δ-Net trained on Matterport3D only

Coverage and Exploration Trajectories



Start: cyan dot · trajectory cyan → pink · grey = unexplored. Classical and NBVP stall; FrontierNet visits room entrances; SCAGE enters each room.

Take-home message

A learned structural prior turns map fidelity into an exploration signal. By resolving geometric anomalies rather than chasing unknown volume, SCAGE achieves higher coverage and lower mapping error simultaneously.

[4] B. Yamauchi, CIRA 1997. [7] L. Schmid et al., RA-L 2020. [8] B. Sun et al., RA-L 2025. [13] X. Wu et al., CVPR 2024. [32] M. Hu et al., TPAMI 2024.

Supported by MeitY, Govt. of India, under "Capacity Building for Human Resource Development in Unmanned Aircraft System (Drone and Related Technology)".